USGS
science for a changing world

NWIS Modernization:
Water Quality Data Management

NWIS Work Program
Joseph Kalfsbeek PMP

11th National Water Monitoring Conference
March 26, 2019

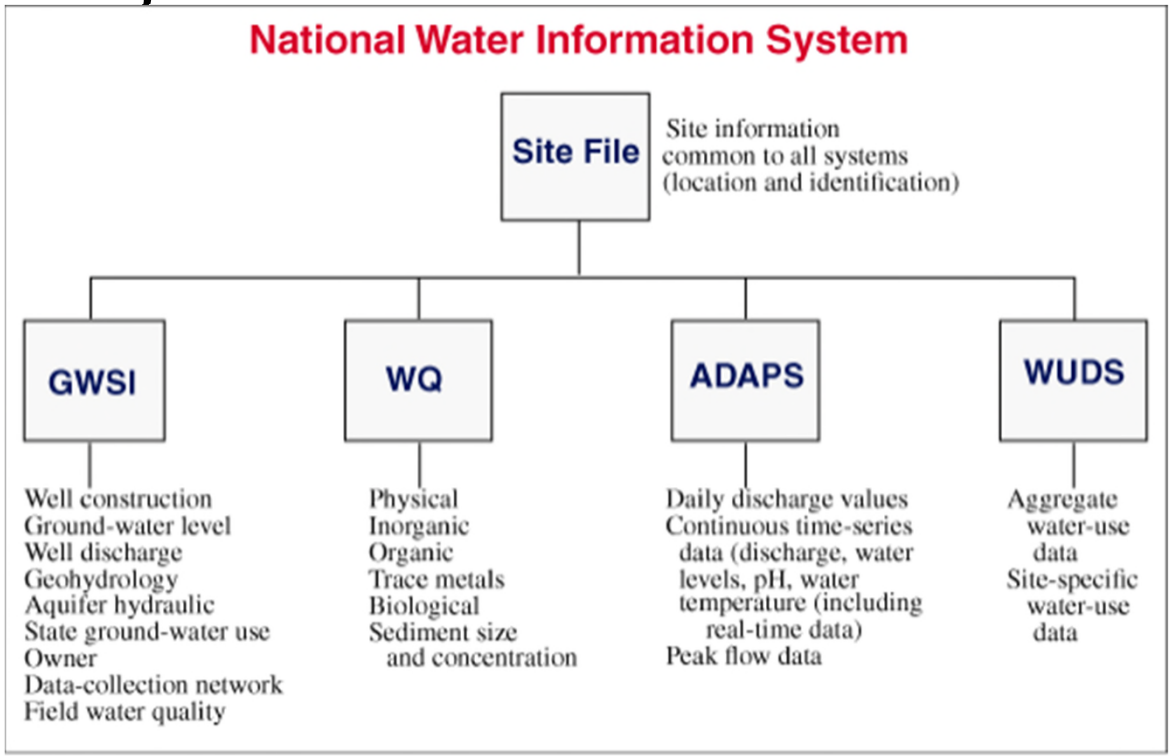
U.S. Department of the Interior
U.S. Geological Survey

Outline

- Background
- Project Information
- Project Goals

USGS
science for a changing world

NWIS Sub-Systems: Pre-Modernization



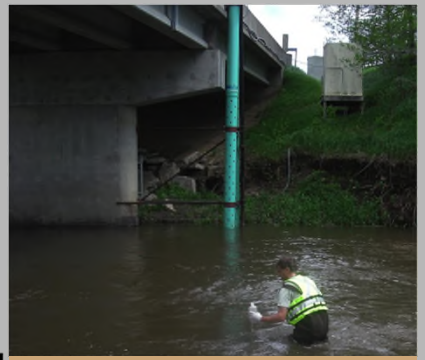
NWIS Components

TIME SERIES

GROUNDWATER

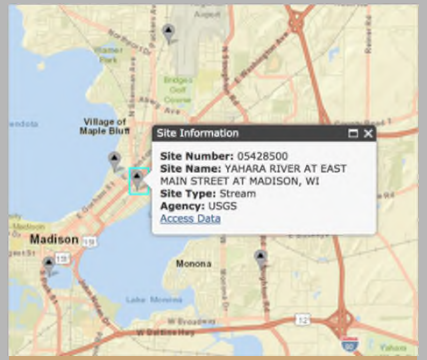
DATA ACQUISITION AND TELEMETRY

FIELD COMPUTING



WATER QUALITY

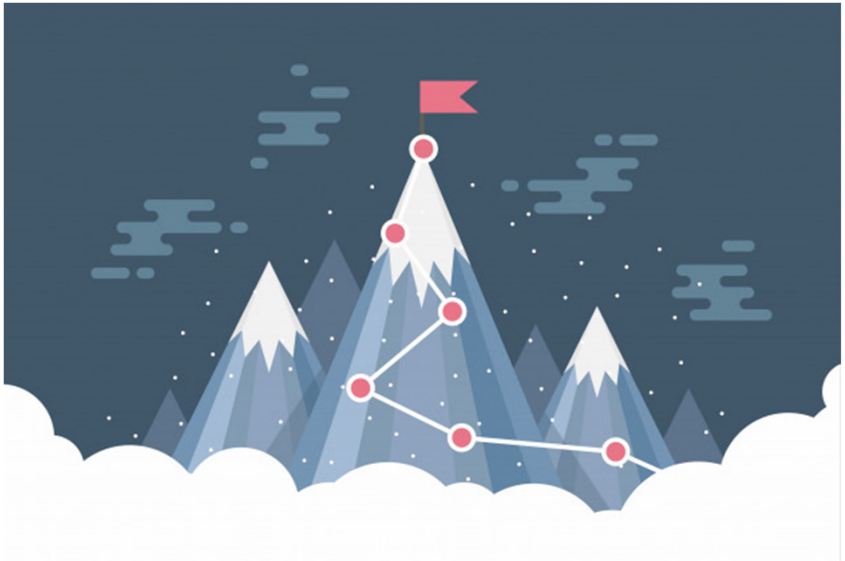
WATER USE



MONITORING LOCATIONS

Reference Lists

The Goal of NWIS Modernization



Our new systems will…

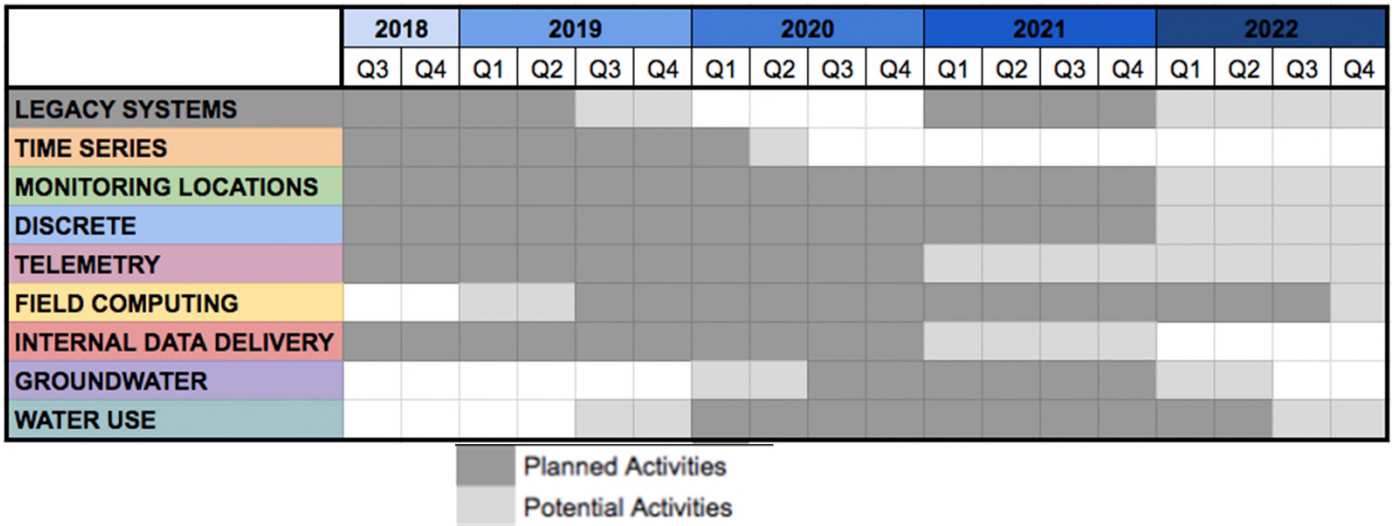
- Support new workflows, collection methods, data types (NGWOS)
- Improve data access and delivery (IOW)
- Harness data from multiple sources (IOW)
- Efficient model assimilation (2WP)

The process of modernizing NWIS…

- Reimplement core capabilities in new systems, either commercial or custom built
- Centralize all systems
- Clean-up current data holdings
- Incremental feature release

NWIS Modernization: Roadmap

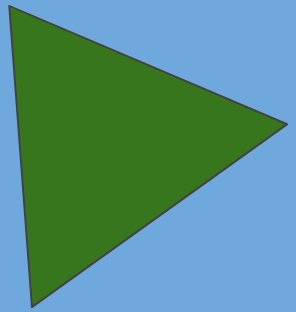
FOCUS AREAS		
TIME SERIES: Curation of continuous and discrete observations over time. Examples: discharge, water levels	**MONITORING LOCATIONS:** Contains record of where data is collected and location metadata	**FIELD COMPUTING:** Tools for data entry during site visits
WATER CHEMISTRY, SEDIMENT & BIOLOGY: Management of discrete sample data and metadata	**GROUNDWATER:** Management of data describing wells, springs, lithology, well construction, aquifers, etc...	**INTERNAL DATA DELIVERY:** Data access and summarization for all data types for internal users
TELEMETRY: Acquisition of time series data.	**WATER USE:** Management of data describing the withdrawal, conveyance, use, and discharge of water	



USGS science for a changing world

AQUARIUS Samples Integration Project

Modernize solution



3 years

$7-$10 Million

USGS
science for a changing world

DISCRETE WATER QUALITY

PROJECT		STATUS		RELEASE DATE	DELIVERABLES
1	WQ/Bio/Sed Market Research	✓	Closed	FY17 Q1	• Market research (phases I & II) • Requirements definition
2	Build vs. Buy Analysis	✓	Closed	FY18 Q4	• Build vs. buy analysis • Alternatives analysis
3	AQUARIUS Samples Integration (ASIP)	L	Planning	FY21 Q4	• Gap analysis • Implement AQUARIUS Samples • Data migration • Improved water quality reports • Training

USGS science for a changing world

Project Scope

- Modernize transactional discrete sample data system
- Support “gage to page” workflow
- Migration of NWIS data
- Seamless integration of discrete and time-series data
- Seamless integration with Field applications, laboratories, public dissemination
- Feasibility of biological data
- Improvements to parameter codes
- Training
- Policy updates
- Close gaps with user needs

Desired timeline for ASIP

Theme	FY19	FY20	FY21	+++
AQ Samples Development				
Infrastructure / Operations				
Integration				
Data Quality/Migration				
Training				
Rollout				

Light | Standard | Heavy

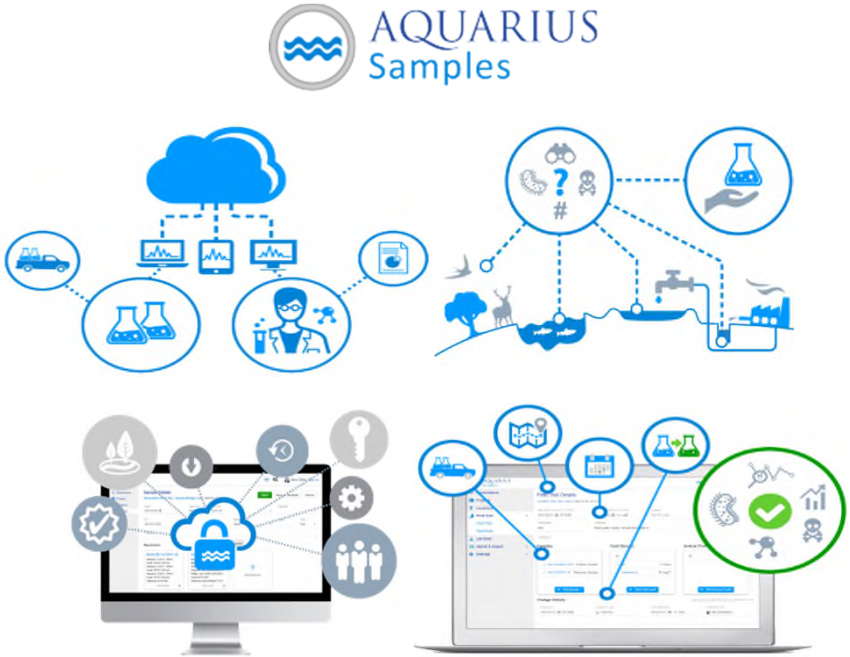
Success

- Widespread adoption
- High user satisfaction
- Seamless Integration
- High confidence in data quality at every point in the workflow
- Users trust the software to grow with their science
- Managers trust the software to grow with their changing vision for WMA science

Project Goals

Goals: Modernize transactional discrete sample data system

- Web-based User Interface
- RESTful API
- Agile
- Flexible
- Scalable
- Hosted in the cloud



Images: Aquatics Informatics

USGS
science for a changing world

Goals: Support “gage to page” workflow

- Provide a single unified ‘standard’ workflow
- Allow flexibility for projects and centers to create their own procedures
- Increase the fidelity of the stored information

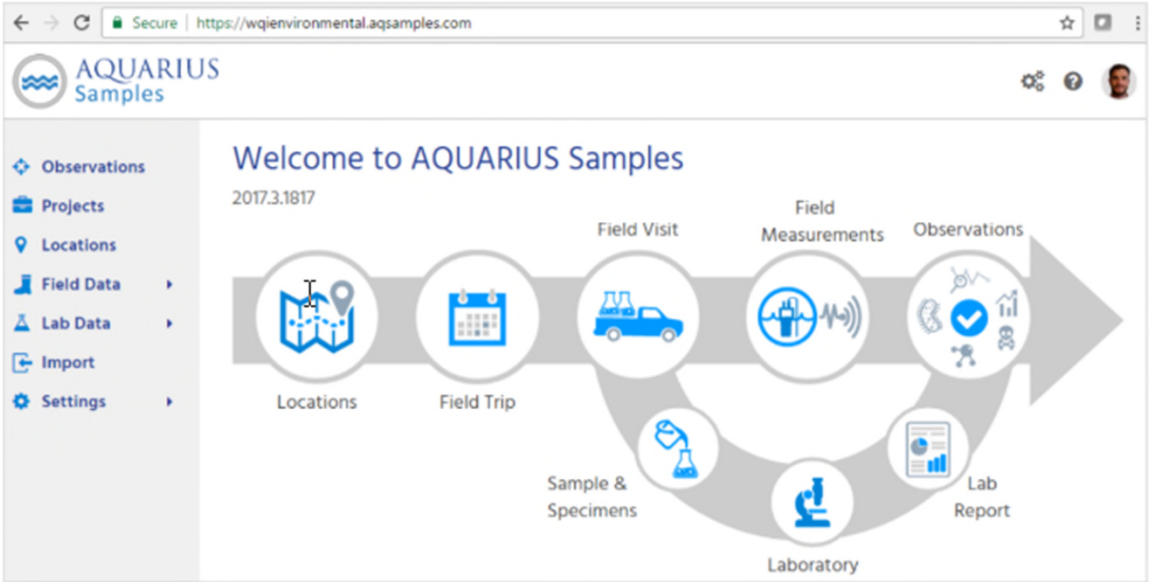


Image: Aquatics Informatics

USGS science for a changing world

Goals: Seamless integration of discrete and time-series data

NWIS Time-Series

Common Tables:

- Locations and site metadata
- Reference lists (parameter, method, medium, etc)

Common Processes:

- Authentication
- Roles and access control

General

Reporting:

- Easy to display discrete sample data on time-series plots
- Easy to compare discrete sample data to closest time-series points in X-Y plots

Workflow:

- Look and feel of data entry and review similar between the two systems

USGS science for a changing world

Goals: Close gaps with user needs

- Sample tracking dashboard
- Auditing of database fields
- Easier control over the display of results, especially around rounding methods or censoring levels
- Ability to visualize data in spatial or temporal context

USGS
science for a changing world

Goals: Close gaps with user needs

- Associate samples in a cross-section, transect, or depth profile
- Associate samples collected for the same project or event
- Associate environmental and quality-control samples
- Store sample collection and handling information at the (subsample) result level

USGS
science for a changing world

Goals: Close gaps with user needs

- Store data and metadata for results from methods that are under development, or that generate multidimensional outputs.
- Store multiple measurements with the same parameter code for the same sample.
- Store photos and other files associated with a sample activity
- Store a Digital Object Identifier with records that comprise a dataset

USGS
science for a changing world

Questions?

Contacts

Sponsor: Cherie Miller (cvmiller@usgs.gov)

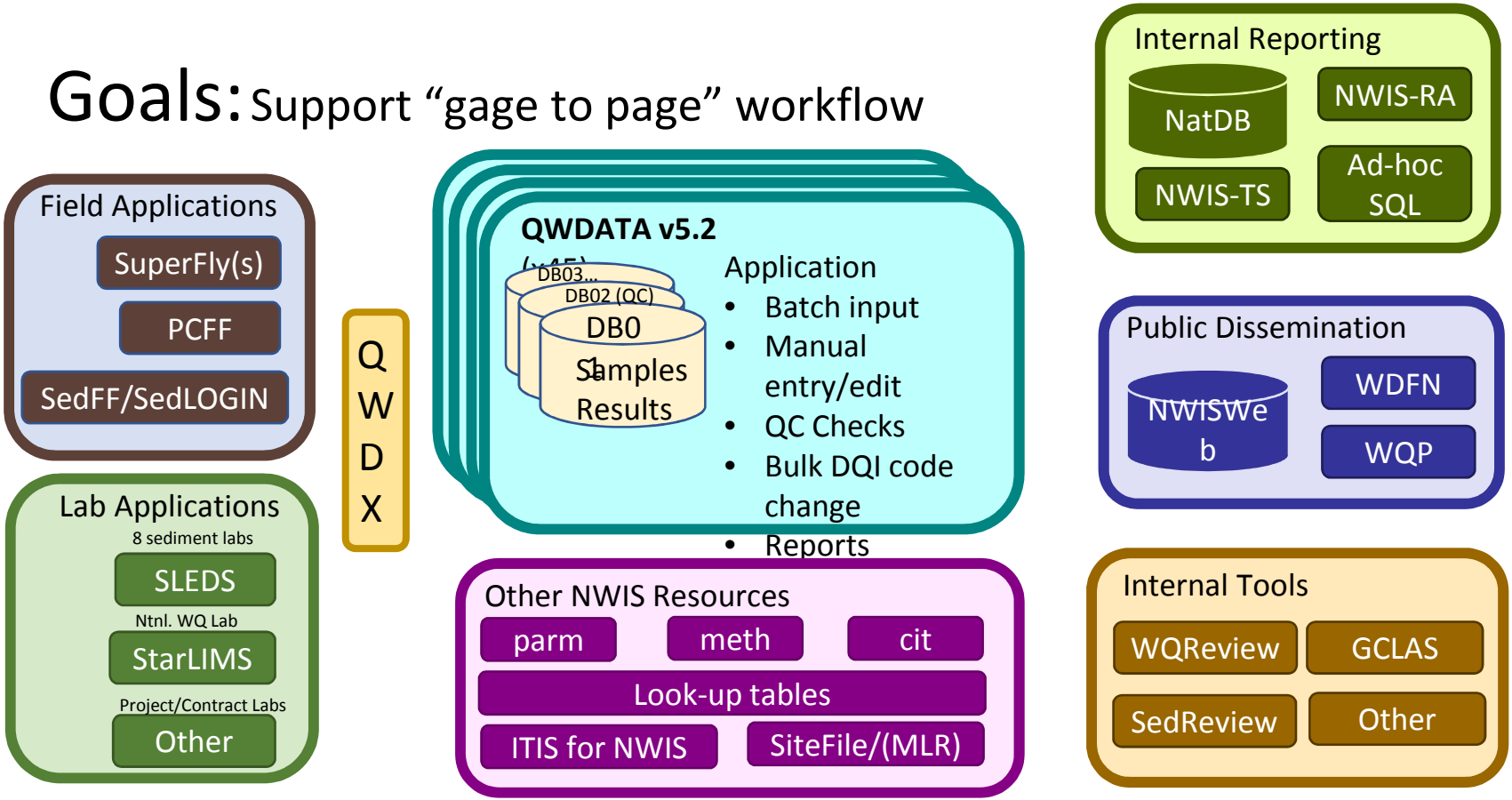
Project Manager: Kate Norton (knorton@usgs.gov)

Deputy Project Manager: Joseph Kalfsbeek (kalfsbek@usgs.gov)

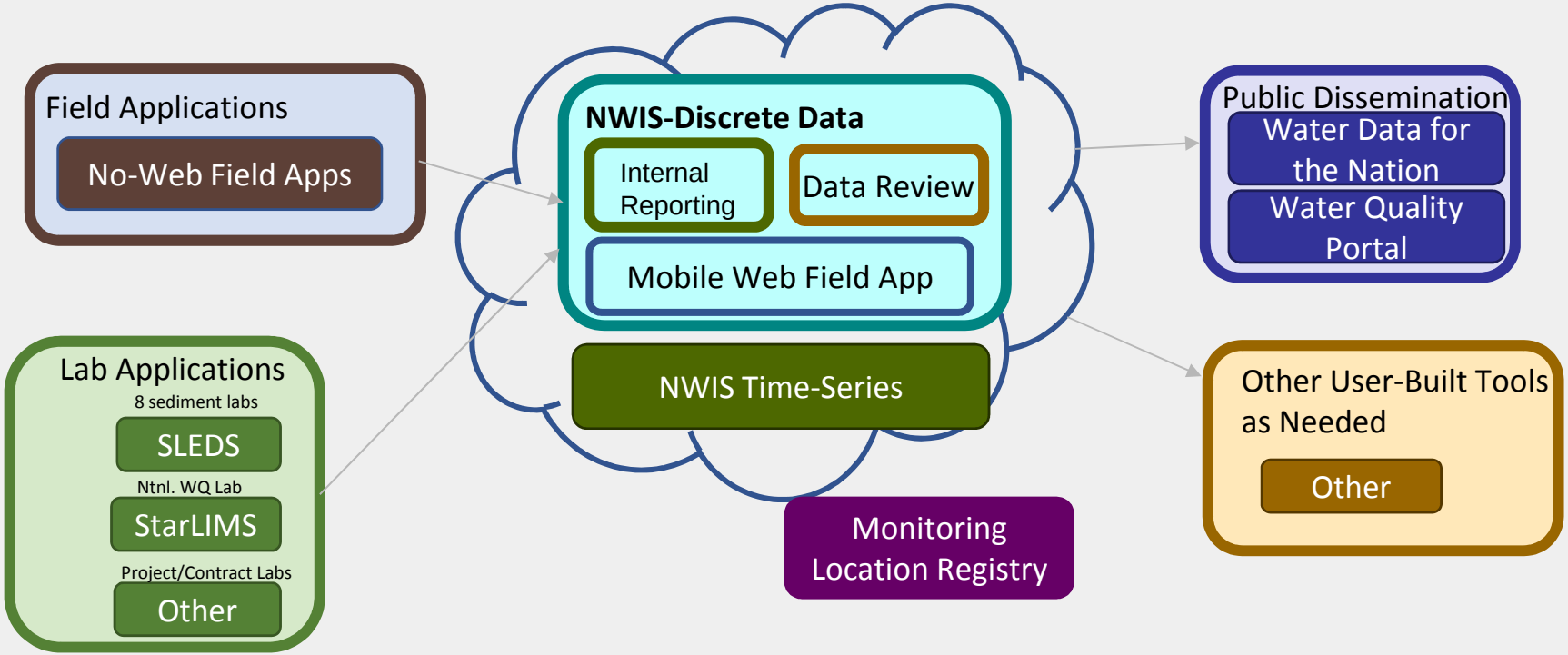
Back Pocket Slides

Architecture

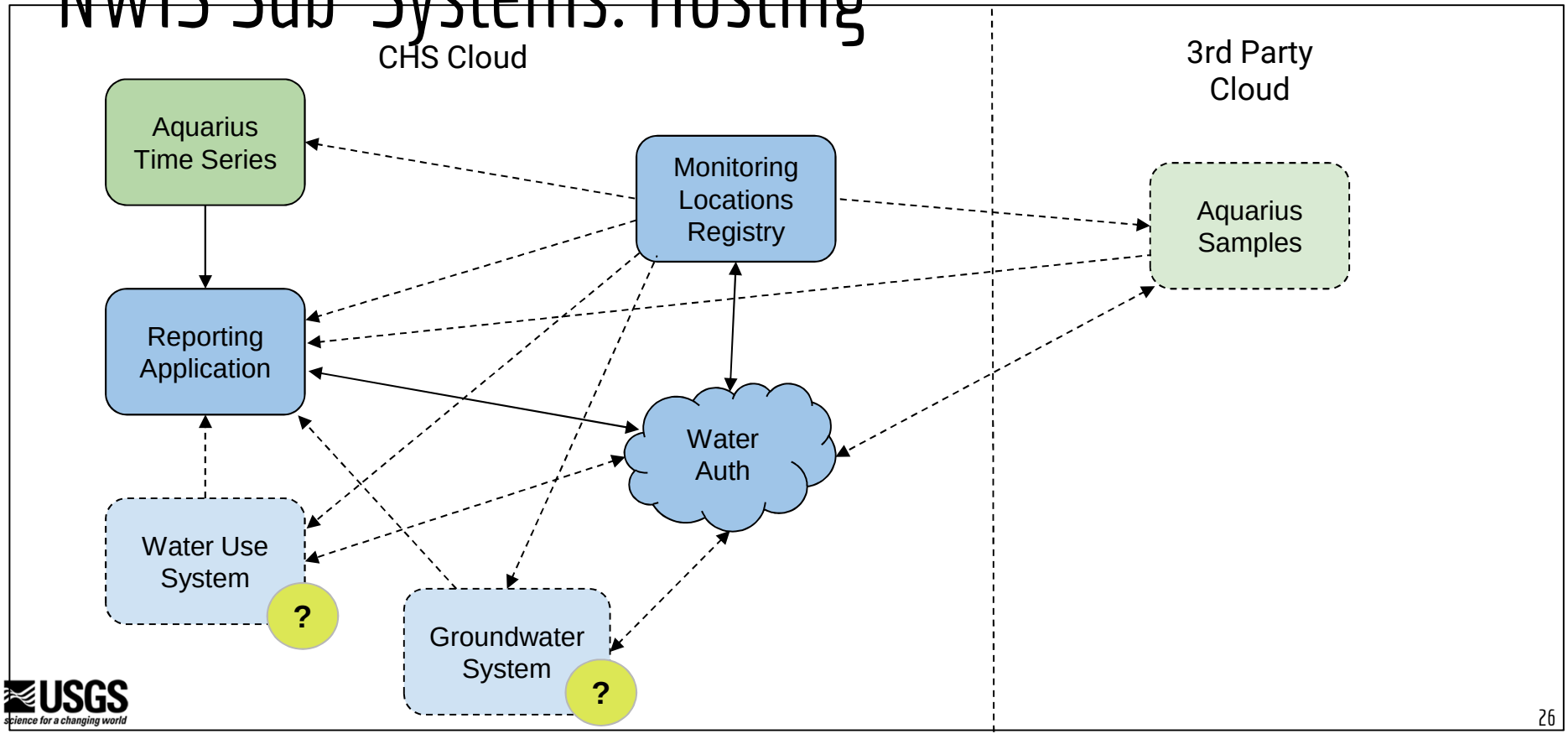
Goals: Support "gage to page" workflow



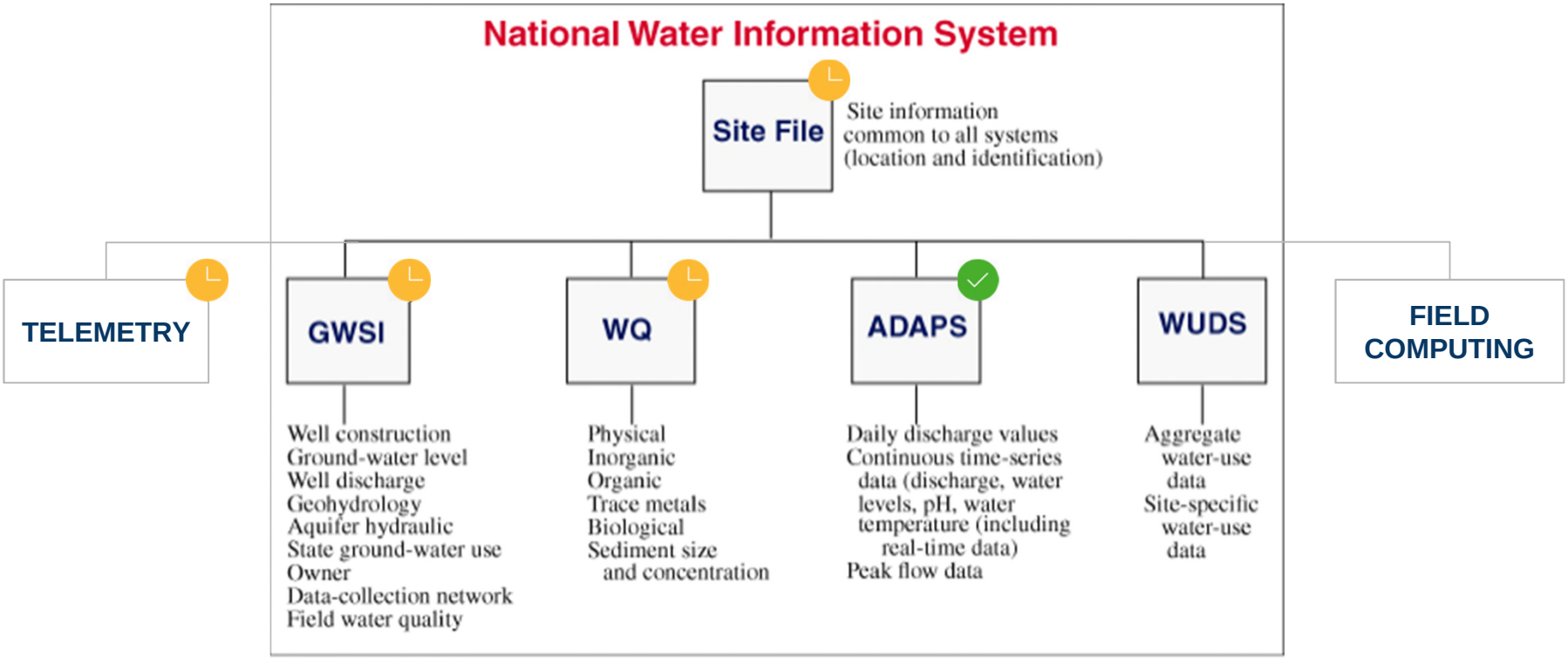
5-year Vision



NWIS Sub-Systems: Hosting



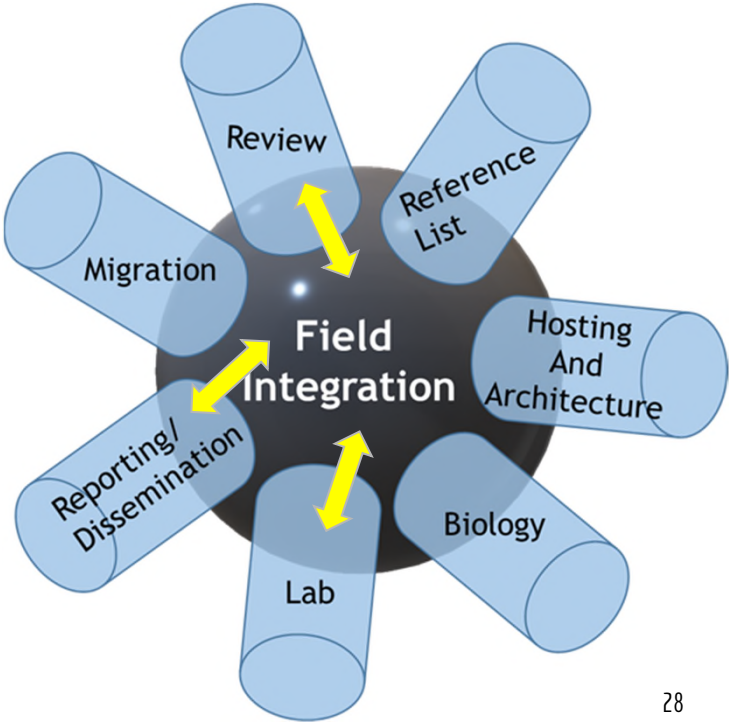
Modernization Status: 2019



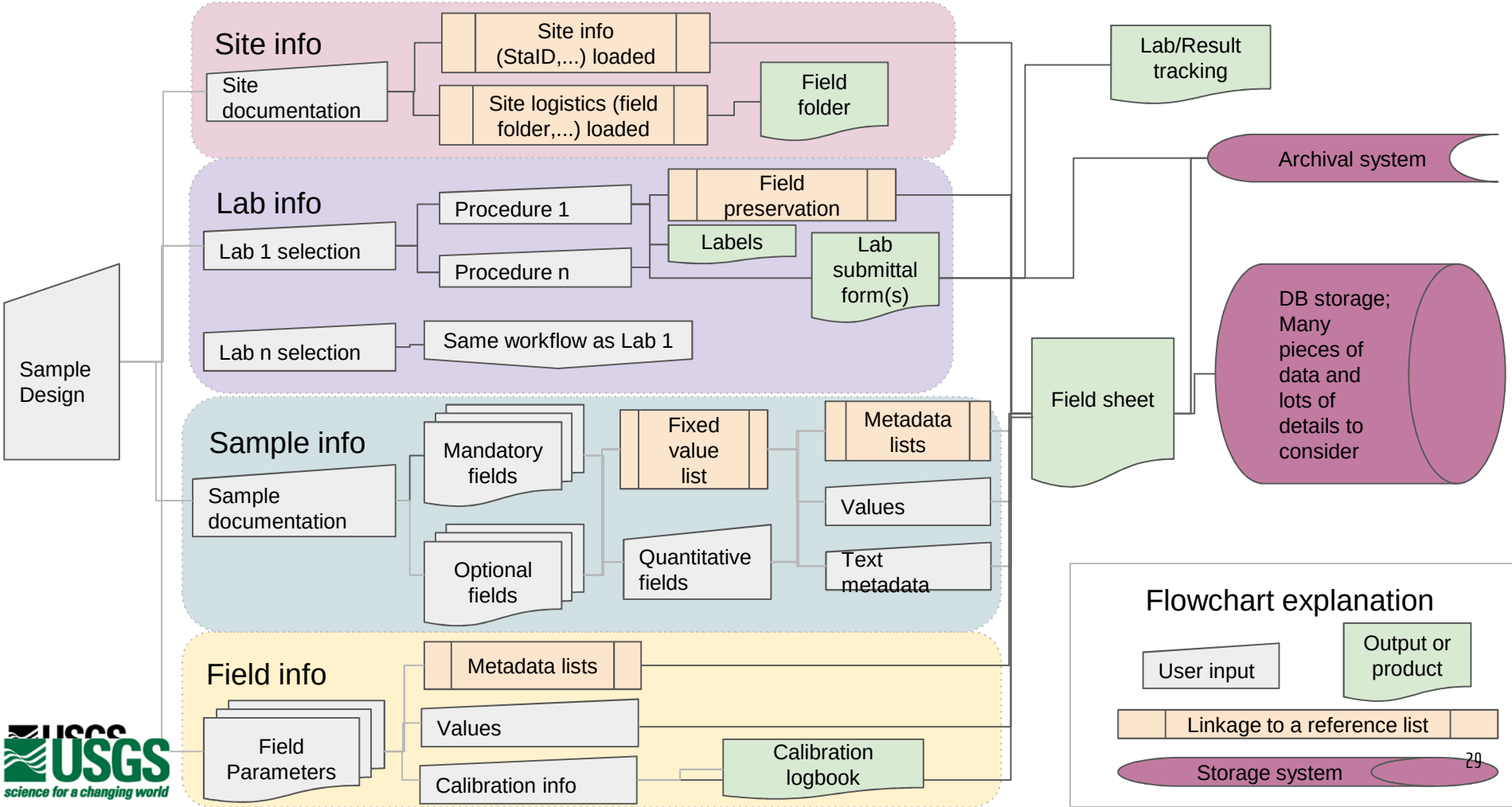
Room for improvement

A dynamic comprehensive Field Report instead of a static field form

- Methods used to determine field parameters
- Notes made during the field visit
- Date samples arrive at laboratory
- Temperature of samples upon arrival at laboratory (if chilled)
- Date sample prepared at laboratory
- Date sample analyses completed at laboratory
- Date sample results received from laboratory
- Date sample data loaded into database
- Lab login comments
- Traceable serial numbers: sondes, probes, Spectrometer
- Connectivity with AQ Time-Series
    - Cross-section observations
- Methods used to determine laboratory parameters
- Laboratory quality-control data



A conceptual flowchart of the USGS field sample workflow



Requirements / User Stories

User Story

As a field data collector, I would like to:

- Know sampling requirements, locations, and needed supplies prior to the field trip
- Follow an efficient and logical workflow for recording measurements
- Ensure all necessary field data are recorded
- Verify data recorded make sense
- Have flexibility to record unplanned conditions or measurements
- Record measurements electronically and upload directly to the database
- Look back over field data in total before uploading to the database

...to ensure field data are complete, accurate, and consistent.

User Story

As a field data collector, I would like for the field form and database to be “smart” enough to warn the user when metadata should have been populated, but is missing. This will prevent missing critical metadata which should have been archived in the database. For example:

- If a sample requires preservation, the preservation lot number and expiration date is required
- If a sample requires filtration, the filter lot number is required
- If a sample is a blank quality-control sample, the blank water lot number, expiration date, and source is required

User Story

As a data collector and hydrologist, I would like the field form and database to be able to record field instrument information and, upon populating this field, require field instrument calibration information. For example:

- Field instrument identification number
- Field instrument type
- Calibration standard type(s) and value(s)
- Calibration standard lot number(s)
- Calibration standard expiration date(s)
- Calibration notes

User Story

As a hydrologist, project manager, and QW Specialist, I would like for the database to be able to track samples from the field, through laboratory analysis, to the population of the database with analytical results. This will aid with:

- Identifying when all data has been received from the laboratory
- Identifying if data are missing, and at what step samples were misplaced or delayed
- Identifying if samples exceeded sample hold times

User Story

As a data collector, hydrologist, project manager, I would like the database to be able to record when an error was corrected in the metadata or field data entry. The database should be able to archive the date of the correction, the person making the correction, and notes about the correction.

User Story

As a hydrologist, I would like the database to be able to record when a laboratory rerun has been requested, by whom and for what reason.

User Story

As a hydrologist, project manager, and QW Specialist, I would like for the database to be able to record and track the date, names, and notes associated with all steps of data review and approval from population of data in the database through data approval. This will aid with:

- Identifying which data still needs to be reviewed and approved
- Identifying the how long data review and approval takes to complete

User Story

As a project chief, WQ specialist, Center Director, or headquarters staff, I would like the data review and approval information to be able to be retrieved in a batch format so that I can see the status and timeline of data review and approval by site, by project, by WSC, or nationally.

User Story

As a hydrologist, project manager, database administrator, I would like the database to complete standard logic checks for QC samples. Often time QC samples are miscoded.
For example:

- If a sample is coded as a blank, the majority of laboratory results should be less than the reporting levels.
- If a sample is coded as a spike, the majority of laboratory results should be detections.

User Story

As a USGS laboratory user, I need to be able to graph laboratory QC data with environmental data from the same laboratory:

- for a specific time period
- for specific analytes or analyte groups
- with the acceptance criteria used by the laboratory
- with user-specified project acceptance criteria
- with the ability to add notes and comments
- saved out electronically to my local server area

so that I can evaluate whether this laboratory can and does meet my project needs. (Required by policy: OWQ Tech Memo 2014.01)

DQI - User Story

As a hydrologist, Project manager, QW specialist, Database administrator I would like the database to be able to record and track the date, names, and notes associated with setting the DQI codes to “Reviewed”.

As a project chief, WQ specialist, Center Director, or headquarters staff, I would like DQI code information to be able to be retrieved in a batch format so that I can see the status and timeline of data approval by site, by project, by WSC, or nationally.

QWVALID - User Story

As a project chief, QW Specialist, or headquarters staff, I would like to be able to complete chemical logic checks, data validation checks, and cation/anion balances on a group of samples at once. I would like the results of the checks to be able to be retrieved in batch format so that I can see the results by site, by project, by WSC, or nationally.

User Story

As a field scientist, I want to be able to efficiently and accurately record sample collection/field metadata while I'm in the field so that I can eliminate transcription errors, save time, and reduce project cost.

User story

As a public user if I find a possible error or problem with water quality data values, I want to be provided with documentation of the USGS data approval process for these data as I assume the USGS applies a minimum, uniform level of quality control for their data, regardless of origin, and especially when tagged as ‘reviewed and approved’. This will help me as a public user to ensure the data is correct and defensible.

USER STORY

As a data collector, project chief, or WQ specialist, I would like to maintain a reporting mechanism similar to WATLISTs so that the User can easily and quickly see what data has been uploaded to the database, can identify and rectify any issues with the sample, and can identify any issues with the results so that samples can be rerun within hold times.

- The reporting mechanism should divide the data by site/project/Center for ease of reviewing.
- The reporting mechanism should provide the data in a format that can be imported to a spreadsheet.

User Stories

As a Laboratory Chemist who uses analytical methods all the time, I need to store information in my customers' database about the analytical methods I use, so my customers and the users of their data will be able to know what methods I used. The details I need to store are are:

- A citation description of the published method, which is usually like a bibliographical entry in the References section of a scientific publication
- The specific method number from the published method, which can contain multiple method numbers in one publication. The NWIS QWDATA database has room to store these citation fields, as well as: (c) a citation name, which serves as a short name for the long citation description; and (d) a method source, which serves as a short identifier for the source agency of the published method (such as USGS, Standard Methods, or

User Stories

As a Field User, I need to properly convey information about the samples I've collected, and what they represent, in a way that is consistent with samples previously collected at my site(s) and within my project. These include the media sampled, and the type of sample relative to its intent, such as a cross-sectional profile, an environmental sample to represent ambient conditions, or QC a sample to assess repeatability or bias, whether from the field equipment or personnel or within the lab.

As a National Researcher, I'd like to be able to select and evaluate results from samples across the Nation that were collected in a similar way, from the same medium, using the same sampling methods, under specific hydrologic conditions.

Why pull the data into Excel?

As a USGS scientist, I want to analyze/interpret water-quality data using Excel. For efficiency, I need a process that retrieves data in a similar format from 1 month to the next (even if I forgot how I did it last month).

In this example, I am a USGS hydrogeologist that wants to represent the ionic characteristics of my project's groundwater sites in 2018 using a Piper Plot. I have an Excel template file that allows me to do this, but I need to import those data.

Important retrieval details: HOST = NWISNE; PROJECT = SVPAS; Dates = 10/1/17 - 10/1/18; Parameters/Properties = Ca, Mg, Na, K, CO3 (alkalinity), Cl, SO4, and Fl;
These are all from public supply wells that are coded as Internal Only and not available online.

USGS
science for a changing world

User Story

- As a USGS Water Quality Specialist, I need to quality assure samples that have been collected and processed by the National Lab. Quality assurance involves a number of ‘sanity’ checks of the different types on constituents tested. Also, possible errors in sample documentation are highlighted for investigation. These checks could be automated or manually retrieved.

Utilities - Inventories of sample information in QWDATA:

- Inventory of WQ samples
- Inventory of DQI codes
- Inventory of method codes
- Inventory of analyzing entities
- Inventory of collection entities

USER STORY

As a project chief, WQ specialist, or national reviewer, I need the database to allow for batch pulls inventorying sample information, DQI codes, method codes, and analyzing and collection entities.

USER STORY

As a WQ specialist, technical reviewer, or database auditor, I would like to generate sample inventory reports to easily determine:

- The number and type of samples and results entered into the database by time, medium, Water Science Center, and/or project
- The number of samples, dates, and parameters for select sites

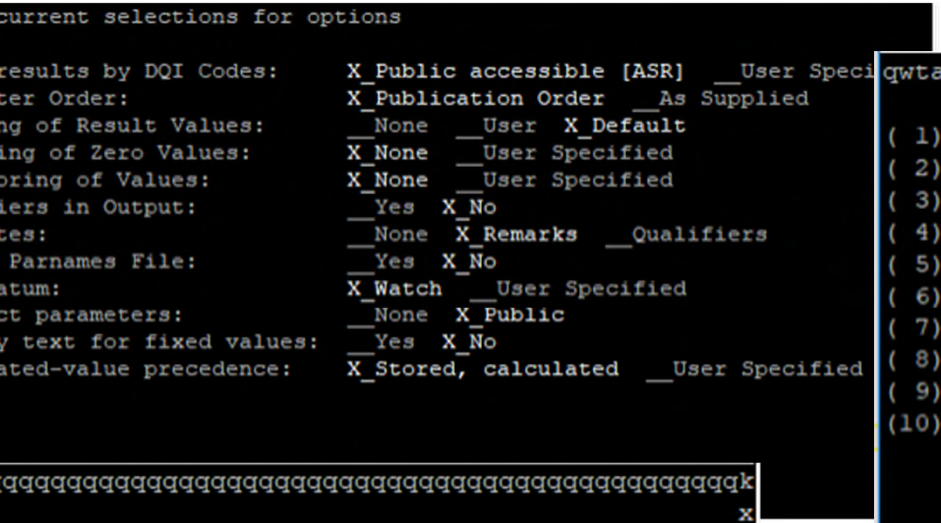
FUTURE USER STORY
As a collector, WQ specialist, technical reviewer, or database auditor, I would like to have missing analytical data readily visible in a summary report for a given sample, sets of samples, projects, or site(s).

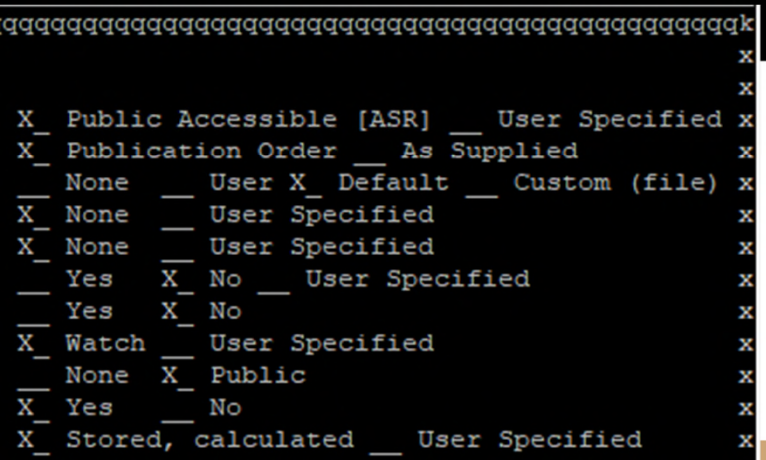
USER STORIES

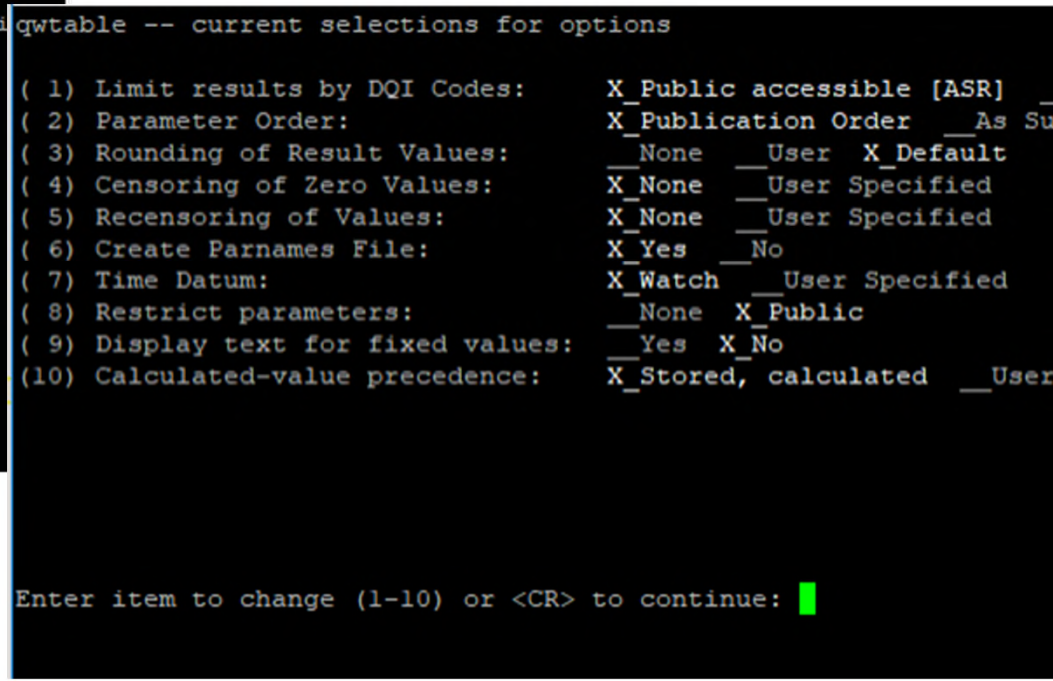
As a WQ specialist, technical reviewer, or database auditor, I would like to generate reports to easily determine the status of data approval (DQI code) by sample, site, time, project, or Center.

As a WQ specialist, technical reviewer, or database auditor, I would like to be able to pull metadata stored with the data approval status, so that I can determine:

- The time taken to approve data
- Who approved or rejected data
- Reasons why data was approved or rejected


current selections for options
results by DQI Codes: X_Public accessible [ASR] __User Speci
ter Order: X_Publication Order __As Supplied
ng of Result Values: __None __User X_Default
ing of Zero Values: X_None __User Specified
oring of Values: X_None __User Specified
iers in Output: __Yes X_No
tes: __None X_Remarks __Qualifiers
Parnames File: __Yes X_No
atum: X_Watch __User Specified
ct parameters: __None X_Public
y text for fixed values: __Yes X_No
ated-value precedence: X_Stored, calculated __User Specified


X_ Public Accessible [ASR] __ User Specified
X_ Publication Order __ As Supplied
__ None __ User X_ Default __ Custom (file)
X_ None __ User Specified
X_ None __ User Specified
__ Yes X_ No __ User Specified
__ Yes X_ No
X_ Watch __ User Specified
__ None X_ Public
X_ Yes __ No
X_ Stored, calculated __ User Specified


qwtable -- current selections for options
(1) Limit results by DQI Codes: X_Public accessible [ASR]
(2) Parameter Order: X_Publication Order __As Su
(3) Rounding of Result Values: __None __User X_Default
(4) Censoring of Zero Values: X_None __User Specified
(5) Recensoring of Values: X_None __User Specified
(6) Create Parnames File: X_Yes __No
(7) Time Datum: X_Watch __User Specified
(8) Restrict parameters: __None X_Public
(9) Display text for fixed values: __Yes X_No
(10) Calculated-value precedence: X_Stored, calculated __User
Enter item to change (1-10) or <CR> to continue: